

BIOLOGIA MOLECULAR: LA NOVA FRONTERA

Per EDUARDO CADENAS

UNIVERSITAT D'ALACANT

MCMLXXXVI

Discurs inaugural

Obertura del curs 1986-1987

“Així que si coneixem quina és la forma dels cossos inanimats o minerals estarem millor capacitats per a seguir la nostra recerca següent sobre les formes dels cossos vegetals i, per fi, la dels animats, que sembla l’etapa superior del coneixement natural del qual és capaç la ment de l’èsser humà”

Robert Hooke, MICROGRAPHIA, 1665

“Un aspecte central de la biologia molecular és la comunicació dinàmica entre els elements genètics, representats pels àcids nucleics, i l’aparell metabòlic, representat pels enzims i altres proteïnes”

H. J. Vogel, 1962

Molt Honorable President de la Generalitat
Excel·lentíssim i Magnífic Rector
Excel·lentíssims i il·lustríssims senyors
Senyores i senyors

Seguint una antiga tradició universitària, avui m'ha correspost l'honor de dictar aquesta lliçó inaugural. El privilegi està basat en l'antiguitat. Constitueix una relíquia d'altres èpoques i altres maneres d'entendre les relacions humanes. Pocs vestigis així aconseguen sobreviure i és possible que aquest també siga sotmès a revisió. També ha sigut tradicional que la lliçó inaugural fóra la lliçó d'introducció a una disciplina concreta. En aquesta lliçó hem triat com a tema parlar-vos de la biologia molecular, la història, la situació actual de frenètic desenvolupament i les perspectives que ens presenta per als pròxims anys en què podria tenir lloc un segon renaixement com a conseqüència de l'aplicació de la biologia molecular a l'ésser humà.

ORIGEN DE LA BIOLOGIA MOLECULAR

Sembla que el terme de biologia molecular va ser encunyat per W. Weaver de la Fundació Rockefeller l'any 1938. Estava preparant un pla de suport a la investigació per a aplicar la ciència física a àrees seleccionades de la biologia, com són la bioquímica, la biologia cel·lular i la genètica. El programa de la fundació va tenir un èxit espectacular. Va ajudar investigadors tan destacats com Rudolf Schoenheimer, Linus Pauling, Robert Robinson i George Beadle, entre d'altres. Però el terme biologia molecular no es va consolidar. Potser va ser perquè va semblar massa presumptuós, però més probablement va ser per la interrupció de l'estudi de les macromolècules, que va implicar l'inici de la Segona Guerra Mundial.

Va ser després que acabara la contesa quan es va popularitzar el terme i se'n perfilà el contingut. Des del primer moment van sorgir dues escoles que se'n van disputar l'hegemonia. G. S. Stent les ha descrites com a escola informacionista i escola estructuralista. La primera, americana, era hostil a la bioquímica; la segona, anglesa, hi estava, en canvi, plenament integrada. Les dues en estreta connexió amb la física, però entenien la relació de formes contraposades. Alguns capdavanters de l'escola informacionista creien en la idea realment fantàstica que la biologia podia proporcionar contribucions significatives al progrés de la física, fins i tot noves lleis i fenòmens. Els pioners de l'escola estructuralista mantenien el punt de vista perfectament raonable que la física podia fer aportacions molt valuoses a la biologia (Stent, 1968).

ESCOLA INFORMACIONISTA

L'escola informacionista de la biologia molecular va tenir l'antecedent en les idees primer de Niels Bohr i després del seu deixeble Max Delbrück. Quan el vitalisme passat de moda estava desapareixent dels cercles intel·lectuals, Bohr va elaborar la noció que alguns fenòmens biològics no es podrien explicar completament en termes de la física convencional. En opinió de Bohr, la dificultat d'entendre la vida en eixos termes radica en el fet que "les condicions que prevalen en la investigació biològica i física no són directament comparables, perquè la necessitat de mantenir l'objecte d'investigació viu imposa restriccions

en la primera per a les quals no hi ha contrapartida en la segona. Així, sembla que hi ha en l'animal viu un *principi d'incertesa* semblant al de l'electró". L'any 1935 Max Delbrück va afirmar que la genètica era el domini de la investigació biològica en què les explicacions físiques i químiques podien resultar insuficients en el sentit expressat per Bohr, de manera que va concloure que "la genètica és autònoma i no s'ha de barrejar amb concepcions fisicoquímiques".

L'interès dels físics en la biologia sembla que, en gran mesura, es va deure a la influència d'un dels líders de la física teòrica. El 1945 va aparèixer un llibre petit titulat *What is life?* escrit per Edwin Schrödinger en què s'anunciava als físics una nova era en la investigació biològica que havia de proporcionar descobriments extraordinaris. El llibre va alçar les passions del seu públic i es va convertir en una mena d'*Uncle Tom's cabin* de la revolució biològica. El llibre va tenir molta influència per l'elegància d'estil i la qualitat intel·lectual. Com deia Crick, "el llibre de Schrödinger va ser molt oportú i va atraure molta gent a la biologia que, d'una altra manera, no s'hi hauria dedicat en absolut" (Crick, 1965). I, a més, va centrar l'atenció sobre el tema del material genètic. Segons Schrödinger, del concepte general de Delbrück de la substància hereditària es dedueix que "la matèria viva, encara que no eludeix les lleis de la física ja establides, és probable que implique altres lleis de la física que, tanmateix, una vegada descobertes, passaran a formar part integral d'eixa ciència com les altres". Watson va llegir també el llibre i, després d'haver-ho fet, va quedar polaritzat l'objectiu de desxifrar el secret del gen (Watson, 1966).

L'estudi dels àcids nucleics descoberts el segle passat va ser impulsat amb força en dues ocasions com a conseqüència del coneixement de la funció biològica. Acabats de descobrir, es van localitzar en el nucli cel·lular (per això el nom). Com a resultat del treball dels citòlegs es va concloure que el material genètic tenia la seu en aquest orgànel cel·lular, de manera que en la dècada dels vuitanta del segle passat ja estava clar que l'ADN havia de ser el material genètic. Com se sap, va ser precisament el progrés en el coneixement de l'estructura, fet en condicions deficientes, allò que conduiria a continuació a l'abandonament d'aquesta noció fins als anys quaranta del nostre segle, que va ser quan es va demostrar definitivament que l'ADN és el material genètic. Els experiments decisius van ser els d'Avery, MacCarthy i MacLeod, que van provar que el factor de transformació és ADN, i els de Hershey i Chase, que van mostrar que el material genètic del fag T2 és el seu ADN. Si l'ADN era el material genètic havia de contenir informació i no podia tenir l'estructura que la hipòtesi tetranucleòtida l'havia assignat. E. Chargaff i el seu grup van fer una extensa determinació de la composició d'ADN de les fonts biològiques més variades i, efectivament, van mostrar variació en la composició segons l'origen de la mostra. A més, van trobar una curiosa regularitat en la composició. El contingut en adenina és igual al contingut en timina, i el contingut en guanina és igual al de citosina (Chargaff, 1950).

Més decisiu va ser encara l'esforç fet per Watson i Crick, els quals, convençuts que l'ADN és el material genètic, van proposar establir-ne l'estructura amb la il·lusió que potser l'estructura, una vegada coneguda, podria revelar com aquesta substància exerceix les diverses funcions que corresponen al material genètic cel·lular (Watson i Crick, 1953).

Van partir de les dades de difracció de raigs X de fibres d'ADN obtingudes per R. Franklin i M. Wilkins del King's College al Regne Unit i van treballar amb models moleculars, com ho havia fet L. Pauling de l'Institut de Tecnologia de Califòrnia per a establir els elements d'estructura secundària de les proteïnes fibroses. L'any 1953 van arribar finalment a la celebrada estructura de la doble hèlix en què l'emparellament de bases complementàries mitjançant enllaços d'hidrogen explicava les regularitats en la composició observades per Chargaff. Efectivament, l'estructura suggeria com du a terme una de les missions assignades al material genètic: la transmissió fidel de la informació de generació en generació. La complementarietat de les bases en les dues cadenes de la doble hèlix els va portar a proposar que en la replicació de l'ADN cada cadena ha de servir de model o motle per a la síntesi d'una nova, amb la qual cosa d'una doble cadena es passa a dues idèntiques entre si i idèntiques a la primera. També constitueix l'estructura de l'ADN el fonament de les altres funcions biològiques estudiades per la genètica. Com va dir Beadle el 1969, "resoldre l'estructura de l'ADN és una de les grans troballes de la biologia del segle XX, comparable en importància als èxits de Darwin i Mendel en el segle XIX, perquè l'estructura de Watson i Crick suggereix d'immediat com es replica o copia en cada generació cel·lular, com s'usa en el desenvolupament i la funció i com experimenta canvis mutables, que són la base de l'evolució orgànica".

ESCOLA ESTRUCTURALISTA

Abans que s'aconseguira provar que l'ADN és el material genètic, les proteïnes havien sigut el candidat generalment acceptat per la importància i la varietat i diversitat de les funcions que exerceixen en la cèl·lula.

La presència universal de la proteïna en els éssers vius va dur Johannes Mulder a proposar en el segle passat el nom que avui tenen i va voler indicar que són de gran importància. Però el seu estudi va adquirir un ímpetu imparable només després que, després d'una llarga polèmica, s'aconseguira demostrar que els enzims són proteïnes. Des d'aleshores aprofundir en el coneixement de l'estructura de les proteïnes era el mateix que fer-ho en la comprensió de com exerceixen la funció aquests catalitzadors biològics, dels quals depenen totes les activitats vitals.

Durant dècades la bioquímica es va ocupar de l'estudi de les transformacions químiques que tenen lloc en les cèl·lules i va descompondre la marxa del procés en etapes individuals mitjançant l'anàlisi de l'estructura dels substrats i productes. Mentrestant, els enzims que fan possible eixes transformacions no s'estudiaven si no era a través de tècniques indirectes i les propietats només eren establides per mitjà d'inferències. Però va arribar un punt en què l'èmfasi va passar a l'estudi de l'estructura detallada de les proteïnes i va ser aleshores quan va nàixer una nova era que s'anomena molecular, però que estrictament s'hauria d'anomenar macromolecular, perquè les proteïnes són macromoleculars.

L'escola estructuralista de la biologia molecular té els orígens en els estudis per difracció de raigs X de proteïnes que es presenten en forma de fibres en què hi ha suficient regularitat. Com deia W. S. Astbury, "les fibres visibles o invisibles són els components estructurals principals de tots els teixits biològics [...] les fibres estan formades per molècules en forma de cadenes i la biologia molecular de moment s'ocupa del partit que la vida ha sabut traure d'aquestes molècules". Astbury va començar la seua vida professional com a físic, va passar a la

cristal·lografia de raigs X i, finalment, a la biologia molecular. La seua entrada en aquesta ciència va estar marcada per la impressió inoblidable que li va causar la regularitat biològica revelada per la difracció de raigs X de les fibres tèxtils tradicionals que són les fibres biològiques.

L'any 1950 Atsbury deia: "sembla que el terme biologia molecular s'està popularitzant bastant i estic content que siga així, perquè, encara que és improbable que siga jo qui l'ha inventat, m'agrada i he tractat de propagar-lo durant molt de temps" (Atsbury, 1950). A continuació assenyalava allò que caracteritza aquesta manera de veure la biologia: "implica no tant una tècnica, sinó més aïna un nou enfocament des del punt de vista de les anomenades ciències bàsiques, amb la intenció de buscar sota les manifestacions a gran escala de la biologia clàssica el pla molecular corresponent". Finalment, indicava clarament l'objecte d'estudi: "s'ocupa particularment de les formes de les molècules biològiques i de l'evolució, explotació i ramificacions d'aquestes formes en l'ascens cap a nivells d'organització cada vegada més alts". I acabava definint-la així: "la biologia molecular és predominantment tridimensional i estructural, però això no vol dir, tanmateix, que siga un mer refinament de la morfologia. Al mateix temps ha d'indagar forçosament sobre gènesi i funció".

Sens cap dubte, aquesta és una excel·lent definició de la biologia molecular per a l'escola estructuralista els antecedents de la qual són les tècniques físiques i l'aplicació a temes químics, en particular la difracció de raigs X aplicada a l'elucidació de l'estructura de molècules de transcendència biològica. W. H. Bragg i son fill W. L. Bragg van inventar la cristal·lografia de raigs X l'any 1912 i després van fundar l'escola de cristal·lògrafs que va convertir el Regne Unit en la pàtria de l'estructura molecular. L'èxit de la tècnica va fer que s'aplicara a molècules cada vegada més complicades. Van ser deixebles seus, com Atsbury i J. D. Bernal, els qui en els anys trenta van començar a investigar l'estructura de les proteïnes i els àcids nucleics convençuts que la funció fisiològica de la cèl·lula es pot arribar a comprendre només en termes de l'estructura tridimensional dels components. Alguns dels seus treballs van proporcionar els primers indicis del camp en què després hi hauria avanços espectaculars. Així, Bernal va aconseguir demostrar el 1939 que el virus del mosaic del tabac, TMV, consisteix en una associació de centenars de subunitats proteíniques idèntiques, i Astbury, el 1945, que en l'ADN les bases formen una pila compacta que és perpendicular a l'eix de la molècula.

Va ser a començament de la dècada dels anys trenta quan Astbury va aconseguir establir l'existència de dos patrons característics en les figures de difracció de les proteïnes fibroses, que va anomenar alfa i beta. No obstant això, la interpretació no va arribar fins als anys cinquanta i no la va proposar un membre de l'escola britànica, sinó Pauling a Califòrnia. Va utilitzar models moleculars desenvolupats i construïts per ell i va fer ús del concepte d'enllaç d'hidrogen que també ell havia desplegat. Així, va poder proposar l'existència dels elements d'estructura secundària, anomenats hèlix alfa i làmina beta, que donaven compte de les pautes de difracció descrites per Astbury.

La interpretació de l'estructura que va fer Pauling indicava una disposició periòdica de la cadena polipeptídica per a generar un entramat extremadament ordenat. La visió, encara que extraordinària des del punt de vista estètic, va comportar una gran decepció ja que no suggeria res sobre com se sintetitzen les proteïnes ni sobre com aconsegueixen exercir les variades funcions biològiques que la cèl·lula els ha conferit (Pauling *et al.*, 1951).

És curiós que els primers avanços espectaculars en el nostre coneixement sobre l'estructura biològica tingueren el començament en l'estudi de les proteïnes fibroses. La raó que utilitzara aquest tipus de material va ser doble. D'una banda, la facilitat de l'obtenció en quantitat sense més guia que les especials propietats fisicoquímiques. D'altra banda, la regularitat de l'estructura de les molècules en si mateixes i de la disposició en la fibra. Però les proteïnes fibroses són una originalitat evolutiva de les cèl·lules eucariòtiques que, sens dubte, tenen molt a veure amb l'habilitat per a construir organismes pluricel·lulars. Es tractava, per tant, de proteïnes poc freqüents. La majoria de les proteïnes són globulars, actuen en dissolució aquosa i estan formades per l'afegitó de subunitats.

L'any 1937 Max Perutz es va proposar a Cambridge establir l'estructura en l'espai de la molècula de la proteïna hemoglobina per difracció de raigs X. En aquest cas no disposava de fibres, sinó de cristalls i el propòsit no era explicar determinades regularitats de l'estructura, sinó obtenir la vertadera disposició, completament desconeguda, dels àtoms en una molècula formada per l'afegitó de quatre subunitats, cadascuna proveïda d'un grup hemo. S'havia escollit l'hemoglobina per ser fàcil d'obtenir i ser relativament petita, com també per la gran importància fisiològica de la funció de transport d'oxigen a la sang i la gran riquesa d'estudis fisicoquímics fets. Semblava probable que l'estructura hauria d'il·luminar la funció en aquest cas com en cap altre. L'any 1959 M. Perutz, juntament amb M. Rossmann, A. F. Cullis, H. Muirhead i T. C. North aconseguien resoldre l'estructura després de 22 anys d'esforç. Quan va ser escomesa l'empresa era impossible de fer, però amb el pas dels anys es van resoldre els problemes teòrics i pràctics i va ser possible coronar-la amb èxit. Abans que Perutz, J. Kendrew aconseguia el 1957 revelar l'estructura d'altra proteïna globular més simple que l'hemoglobina, la mioglobina, que en el múscul, especialment de mamífers marins, emmagatzema oxigen per a suplir les necessitats respiratòries (Dickerson i Geis, 1981).

Potser l'aspecte més curiós de les dues estructures va ser que incloguera nombroses hèlixs alfa amb una estructura exactament coincident amb la que Pauling havia proposat per a les proteïnes fibroses i es va basar en els models moleculars. Però les hèlixs no estaven disposades de cap manera en una estructura regular i la impressió que causava la molècula, en paraules de Perutz, era visceral i repugnant. Una vegada més, l'estructura resultava decebedora perquè no semblava il·luminar com l'hemoglobina exerceix la seua funció. La cooperativitat en la unió amb l'oxigen, coneguda des de començament de segle i atribuïda a una interacció directa entre els grups hemo, quedava descartada per l'allunyament d'aquests grups en la molècula.

Coincidint en el temps en bona mesura en el projecte de determinació de l'estructura de les primeres proteïnes globulars, es va convertir en realitat en altre projecte brillant. Va ser el treball de Frederick Sanger i el seu grup que es proposava la seqüenciació d'una proteïna. Quan el va dur a terme semblava raonable suposar que totes les molècules d'una proteïna tingueren els mateixos residus d'aminoàcid i que aquests aminoàcids estigueren disposats en el mateix ordre. Però aquest punt de vista, que sostenia Sanger, no era compartit per tots els investigadors, molts dels quals pensaven que, si era cert, comportaria una enorme càrrega per a la cèl·lula. Sanger va fer un treball enormement original. Va preparar l'estratègia de la seqüenciació, posar a punt les tècniques necessàries, algunes de les quals va haver de desplegar ell mateix i, finalment, fer el seu propòsit a través d'un esforç continuat de quasi una dècada. Va començar el projecte l'any 1944. Va elegir l'hormona insulina perquè era relativament petita: només té 51 residus i n'hi havia en quantitat. A més, es pensava que el

coneixement de la seqüència potser revelara com actua per provocar les nombroses accions cel·lulars, que contribueixen eficaçment a mantenir el nivell de glucosa a la sang, dins els límits adequats per a l'organisme animal. Va aconseguir desenvolupar un mètode molt convenient per a marcar residus N terminals, que va ser decisiu per a l'èxit del projecte. Finalment, l'any 1953 Sanger i Thomson ja havien completat la seqüència de les cadenes A i B de l'hormona. Dos anys després van aconseguir establir la posició dels ponts disulfur en la molècula de la insulina. Així havien provat que totes les proteïnes tenen seqüència i per això contenen informació i també acabaven de provar per primera vegada que una cadena polipeptídica no conté altre tipus d'enllaços que els enllaços peptídics.

Una vegada desenvolupades les tècniques que permetien seqüenciar una proteïna i establir-ne l'estructura en l'espai, l'aplicació va permetre conèixer-ne amb detall moltes altres. Però, al començament, el coneixement de l'estructura no va obrir les portes a la comprensió de com les proteïnes exerceixen les accions que les caracteritzen. Prompte es va proposar que el problema era que la molècula proteínica és molt gran i, en bona mesura, l'estructura no és rellevant a l'acció. Només ha de ser-ho una part que és necessari identificar. Amb el propòsit de disposar dels criteris que ho feren possible es va iniciar una fèrtil àrea d'investigació en què, a través de la comparació de la mateixa proteïna en espècies diferents, s'esperava trobar les regions conservades de la molècula i, en aquestes, la base de l'acció biològica. Efectivament, es van poder detectar les regions importants de la molècula com les regions més conservades, però no va ser possible interpretar per què eren importants. Però aquest fracàs va ser compensat amb escreix quan les dades obtingudes a través d'aquest enfocament comparat van ser utilitzades per a investigar la marxa que havia seguit l'evolució d'una mateixa proteïna en diferents espècies. Així es va poder aprofundir, com mai no s'havia pogut fer, en els mecanismes a través dels quals ha tingut lloc l'evolució molecular. A més, les dades convenientment elaborades van servir per a establir arbres filogenètics que, en el cas dels organismes superiors, es va poder comprovar que coincidien molt bé amb els obtinguts utilitzant caràcters morfològics. Amb aquesta confiança, el mateix enfocament molecular es va poder aplicar als bacteris en què els criteris morfològics no havien sigut capaços d'elaborar arbres filogenètics adequats.

Finalment, els intents de comprendre la funció a través de l'estructura van tenir èxit amb els enzims. No molt de temps després que s'establira les de mioglobina i hemoglobina, es va determinar l'estructura tridimensional del primer enzim, el lisozim. Ja amb aquest enzim va ser possible aclarir nombrosos detalls sobre el mecanisme de la reacció que catalitza. I el mateix ha passat des d'aleshores amb gran varietat d'enzims. Però, per a aconseguir-ho, altra vegada va ser necessari utilitzar l'enfocament comparatiu, només que ara era necessari comparar l'estructura en l'espai de l'enzim lliure amb l'estructura amb lligands adequats units als seus centres perifèrics. D'aquesta manera, en els casos millor estudiats hi ha un conjunt d'estructures que, a manera de vinyetes, descriuen tot el mecanisme de la reacció catalitzada amb detall atòmic. I allò més interessant és que els resultats així obtinguts van confirmar els mecanismes proposats a partir de dades de caràcter indirecte reunides després de molts anys de tasca pacient dels enzimòlegs.

Altra tasca de gran interès de l'enzimologia ha sigut l'estudi de la regulació de l'activitat enzimàtica. Les primeres troballes van tenir lloc quan, en els anys cinquanta, es va iniciar per fi l'estudi de les vies anabòliques que fins aleshores havien resistit els intents fets per a poder aclarir-les. Aleshores el progrés va ser,

en canvi, molt ràpid. La raó que va justificar el ràpid avanç va ser l'ús de l'anàlisi genètica i, alhora, els mètodes clàssics de la bioquímica. Va ser en aquella època en què es van fer les primeres observacions decisives sobre els dispositius reguladors i quan, a través del fenomen de la inhibició pel producte final de la via, es va reconèixer l'existència dels enzims reguladors. Va ser també aleshores quan es van perfilar les primeres observacions sobre el control de l'activitat enzimàtica per fosforilació i desfosforilació en el cas de la fosforilasa del glucogen. Amb els pas dels anys, la regulació al·lostèrica i la regulació per interconversió arribarien a ser reconegudes com dos mecanismes reguladors bàsics i d'àmplia distribució en la natura. D'altra banda, la regulació hormonal va tractar el problema de la naturalesa dels efectors i en eixe camp es van establir interessants connexions entre la interconversió enzimàtica com a dispositiu generador de la resposta i la unió de la molècula activa amb el receptor com a estímul. Va ser també a final dels cinquanta quan es va reconèixer l'existència del camp com a segon missatger. També en aquells anys Monod desenvolupava la intensa tasca que el va conduir a començament dels anys seixanta a proposar el model de l'operó per a explicar la regulació de la síntesi de proteïnes en els bacteris i, a mitjan la mateixa dècada, el primer model al·lostèric que comprenia com un element essencial els canvis de conformació en la proteïna reguladora (Cadenas, 1978).

ACCIÓ DELS GENS

Algunes de les funcions assignades al material genètic havien pogut ser explicades amb referència a l'estructura de la doble hèlix de l'ADN i la complementarietat de les bases en una cadena i altra de la molècula. No obstant això, calia una descripció molecular de com els gens governen les activitats de cada dia a les cèl·lules que els allotgen. La història de com es va aconseguir aquest objectiu constitueix un dels capítols més brillants de la biologia molecular escrit pels investigadors més extraordinaris que ha tingut aquesta ciència.

La síntesi de proteïnes està, en últim terme, governada per l'ADN. El procés és molt complex. Hi intervé un gran nombre de components. El control per part de l'ADN també s'exerceix d'una manera complicada. Per això, l'etapa inicial en l'elucidació de la síntesi de proteïnes sota el control genètic va estar determinada per esdeveniments que consistien en l'establiment de connexions que comportaven la implicació de diferents substàncies en el procés.

Hi ha una afecció artrítica als pacients de la qual presenten el color del vi en l'orina. És l'alcaptonúria, una malaltia hereditària. L'any 1902 Sir Archibald Garrod va deduir d'un estudi dels antecedents familiars dels pacients que presentaven aquest desajust que es transmet per un gen recessiu. Es tracta d'una mutació que determina un defecte metabòlic. La substància colorejada normalment es transforma, però en aquests pacients no és així i s'excreta de manera ostensible en l'orina. Garrod va reunir diversos exemples d'aquest tipus i va suggerir per a descriure'ls el terme *inborn error of metabolism*. Així era assenyalada per primera vegada una connexió irrefutable entre gens i enzims. Però les idees eren molt avançades per a l'època i el concepte que els gens actuen directament sobre el metabolisme no va tenir àmplia acceptació fins als anys quaranta gràcies als experiments amb mutants bioquímics de *Neurospora crassa* fets per G. W. Beadle i E. L. Tatum.

Com a resultat i resum de la interpretació d'aquests treballs es va crear l'aforisme un gen = un enzim. Amb aquest aforisme es volia expressar que la

funció primària de cada gen consisteix a dirigir la síntesi d'un enzim i, d'aquesta manera, controlar una reacció metabòlica concreta.

Hi ha una anèmia que preval en certes regions de l'Àfrica. S'anomena anèmia de cèl·lules falciforme perquè els glòbuls rojos dels pacients d'aquesta greu malaltia adopten una forma de falç molt característica. L'any 1949 Pauling, juntament amb H. Itano i S. J. Singer, van provar que aquesta anèmia es deu a la possessió d'una hemoglobina defectuosa que té una càrrega elèctrica menor de la normal com s'evidencia a través de l'electroforesi. Van descriure el defecte com a malaltia molecular. Set anys més tard, V. M. Ingram aconseguia provar que la molècula de l'hemoglobina en l'anèmia de cèl·lules falciforme difereix de la normal solament en la substitució d'un residu de l'aminoàcid glutamat pel de valina. Mai abans no s'havia provat que una mutació consistira en la substitució d'un residu per un altre en una cadena polipeptídica. Així, hi havia un indici que indicava com un gen aconseguia controlar la síntesi d'un enzim. Seguint-lo es va arribar finalment a la conclusió que allò que fa el gen és determinar la seqüència de la proteïna, la síntesi del qual governa. Però l'ADN no ho fa directament, requereix el concurs de diferents tipus d'ARN, l'altra classe d'àcid nucleic. Van ser principalment els treballs de J. Brachet i de T. Caspersson els que a mitjan anys quaranta i com a resultat de l'aplicació de les tècniques histoquímiques van assenyalar que les cèl·lules implicades activament en la síntesi de proteïnes tenen un alt contingut en ARN. Així, els investigadors van començar a familiaritzar-se amb la implicació de l'ARN en la síntesi de proteïnes. Però abans que es poguera arribar a un coneixement detallat sobre el funcionament d'això paper es va necessitar disposar de sistemes *in vitro* per a la síntesi de proteïnes.

Zamecnik i el seu grup van obtenir a Boston l'any 1954 el primer sistema desprovisat de cèl·lules capaç d'incorporar *in vitro* aminoàcids en una cadena lineal a través de la formació d'enllaços peptídics. El mateix sistema va revelar la participació dels ARN de transferència, la necessitat de GTP i que la síntesi té lloc en els ribosomes rics en ARN, però no va poder mostrar l'existència de l'ARN missatger, que du la informació genètica des de l'ADN als ribosomes. Van ser precisament els experiments de Jacob i Monod sobre la regulació de la síntesi de proteïnes en els bacteris els que van posar de manifest la necessitat de l'existència d'un missatger i van revelar-ne les propietats bàsiques. L'evidència era aleshores indirecta, però poc després els experiments d'E. Volkin i L. Astrachan primer i els de Spiegelman després provarien l'any 1980 l'existència de l'ARN i mostrarien, a través de la comparació de la composició els uns i de la hibridació els altres, que molt probablement el missatger tenia una seqüència que era transcripció de la seqüència de l'ADN, tal com s'havia postulat. L'últim pas, el de la mateixa acció biològica, va ser fet finalment per M. Nirenberg i J. Matthaei, que van exposar en el Congrés Internacional de Bioquímica de 1961 a Moscou que missatgers naturals i artificials estimulaven la síntesi de proteïnes en un sistema *in vitro* (Nirenberg i Matthaei, 1961).

El mateix sistema descrit per Nirenberg i Matthaei va servir al seu grup i al de S. Ochoa per a establir les bases de la correspondència entre triplets de bases en el missatger i aminoàcids incorporats a la proteïna. Aquesta correspondència descrita com a codi genètic va ser obtinguda a través d'enginyosos procediments, que es van avançar a la realització mitjançant la seqüenciació d'una proteïna i del gen que la codifica, que hauria sigut el mètode més directe però aleshores més difícil. El progrés va ser ràpid i prompte es va revelar amb tots els pèls i senyals el codi, que va ser àmpliament aclamat com a dipositari del secret de la vida.

Des que es desenvoluparen els primers sistemes *in vitro* per a la síntesi de proteïnes hi va haver un progrés ininterromput en el coneixement, no solament de la implicació de més i més components en el procés, sinó de més detalls moleculars pel que fa al paper que juguen en el procés en el conjunt d'acord amb un coneixement també exhaustiu de l'estructura. El progrés va afectar primer els ARN de transferència que van ser les primeres molècules d'àcid nucleic seqüenciades. També va ser un ARN de transferència el primer ARN l'estructura del qual en l'espai va ser establida a través de difracció de raigs X. Més tard, l'interès es va traslladar als ribosomes i els seus components, les proteïnes ribosòmiques i els ARN ribosòmics. Es van descobrir factors proteïnics implicats en el procés i es va aconseguir establir el paper que juga el GTP. Finalment, i en un progrés encara no conclòs, s'està intentant tenir una imatge com més detallada millor de l'organització dels components al ribosoma, les interaccions d'aquest en les diverses etapes de la síntesi amb els diferents factors i els mecanismes moleculars dels quals depèn la catàlisi, com també la fidelitat del procés global de síntesi.

L'esforç fet per a conèixer els components de l'aparell de síntesi de proteïnes va tenir un fruit col·lateral d'una gran importància que hauria de canviar ni més ni menys que les nostres idees sobre l'origen i l'evolució de totes les cèl·lules al nostre planeta. El 1965, gràcies als esforços de molts investigadors i en particular de Zuckerkandl i Pauling, ja s'havia comprès que les macromolècules tanquen un registre de la història evolutiva de la vida. Aquest resultava especialment pertinent per als bacteris l'evolució dels quals no era possible desxifrar a través de l'ús de la morfologia comparada. Woese i els seus col·laboradors van iniciar un projecte per a aclarir la filogènia d'aquest grup tan antic d'organismes mitjançant la comparació de les macromolècules de les diferents espècies. Tanmateix, no van tractar la comparació de les seqüències d'alguna proteïna escollida. L'any 1972 van començar a estudiar ARN ribosòmics convençuts que havien de ser millors que les proteïnes perquè semblen evolucionar més espai i no alteren en res el paper fisiològic que els va ser assignat originalment. Com que aleshores no resultava pràctic seqüenciar fins el més petit dels ARN ribosòmics, van optar per elaborar catàlegs del més petit dels ARN ribosòmics grans. Per a elaborar els catàlegs només es fragmentava l'ARN i se seqüenciaven els fragments, però no s'intentava ordenar aquests fragments en la seqüència total. El projecte es va dur a terme amb una gran varietat d'espècies bacterianes i algunes eucariòtiques. Però els resultats totalment inesperats van forçar una revisió dels conceptes sobre l'evolució en els estadis més primerencs, que encara no han acabat de completar-se però que ja ha capgirat totes les idees que prevalien amb anterioritat.

L'any 1977 Woese i els seus col·laboradors van proposar l'existència de dos grups diferents de bacteris que no estan emparentats entre si més estretament del que ho puguem estar amb les cèl·lules eucariòtiques. Un dels grups va ser anomenat eubacteris i comprenia la major part dels bacteris coneguts. L'altre va ser batejat amb el nom d'arqueobacteris, que originalment només englobava els bacteris metanògens. El 1978 es va mostrar que els organismes halòfils extrems de les salines són també arqueobacteris i més tard es va provar el mateix per a un conjunt d'organismes que solen viure a altes temperatures i tenen un metabolisme que depèn del sofre i que comparteixen amb la resta d'arqueobacteris la possessió de lípids molt poc usuals. El nom d'arqueobacteri suggereix antiguitat. Les propietats semblen que són les que degueren tenir els organismes que van viure en l'era arcaica i que prosperaven en condicions que ara ens semblen extremes (Woese i Fox, 1977; Woese *et al.*, 1978; Fox *et al.*, 1980).

ENGINYERIA DELS GENS

La microbiologia i els seus mètodes van tenir una gran transcendència en el creixement de la biologia molecular. Ja hem esmentat l'escola americana informacionista i M. Delbrück, que va formar el famós grup dels fags i va protagonitzar una època de progrés espectacular. També van ser bacterians els sistemes *in vitro* que van servir per a revelar la naturalesa del codi genètic. Van ser també estudis amb bacteris els que van dur Jacob i Monod al seu model de l'operó. En moltes d'aquestes investigacions les tècniques van tenir originalment caràcter genètic per a passar després a ser sotmeses a l'escrutini dels mètodes bioquímics. Des dels anys seixanta el progrés ha sigut continuat en el nostre coneixement quant a les manipulacions a què és sotmès l'ADN a la cèl·lula. No solament s'ha aconseguit així un gran avanç en els coneixements, sinó també un enorme progrés en les tècniques de maneig, com també en l'arsenal d'enzims disponibles per a fer-ho, fins aconseguir que l'ADN siga la substància amb què es pot dur a terme el major nombre d'alteracions deliberades. La conjunció de totes aquestes tècniques va provocar en la dècada dels setanta el naixement de tota una nova tecnologia que ha rebut diverses denominacions. Es parla d'enginyeria genètica perquè permet alterar la voluntat de la riquesa genètica dels organismes per al nostre profit. S'anomena tecnologia de l'ADN recombinant perquè permet la preparació de molècules d'ADN mixtes que, a manera de quimeres, tenen porcions que provenen d'espècies diferents. Ens hem acostumat al clonatge molecular que permet l'obtenció de qualsevol fragment d'ADN en què estiguem interessats i en qualsevol quantitat que considerem necessària. Aquesta tecnologia permet canviar tots els aspectes de la nostra vida en les aplicacions pràctiques, però de moment allò més important que està fent és permetre l'avanç de la biologia molecular en tots els camps, fins i tot en aquelles àrees que, com el desenvolupament o el cervell, s'havien resistit més als atacs.

El clonatge molecular permet obtenir en quantitat qualsevol gen, tant procariòtic com eucariòtic, i després el desenvolupament de mètodes avançats per Gilbert i Maxam i per Sanger fan possible i relativament senzilla la seqüenciació de l'ADN. És aquesta tècnica la que revoluciona tants camps d'investigació. No solament fa factible la seqüenciació d'una proteïna, seqüenciat el gen que la codifica. Permet fer el mateix per a proteïnes totalment desconegudes que encara no han sigut aïllades i, per això, no se sap quina funció poden exercir. A més, facilita no solament la seqüència d'una proteïna, sinó la del seu precursor i dóna indicacions sobre el camí que segueix la seua elaboració després de la síntesi. Permet conèixer la seqüència i, fins i tot, inferir l'estructura de proteïnes de membrana, com ara receptors, o canals. Fa possible acumular grans quantitats de seqüències proteíniques que es poden comparar i descobrir analogies insospitades que indiquen connexions evolutives molt interessants. En definitiva, ha canviat la nova tecnologia els laboratoris, la naturalesa de les publicacions científiques i promet una renovació total de la biologia molecular, que per a aprofundir en els mecanismes enzimàtics pot aprofitar les tècniques de mutació dirigida que li permeten preparar proteïnes amb l'estructura alterada, establir la importància de qualsevol aspecte cultural que ens proposem assajar i, potser, arribar així a la fabricació de superenzims; per a comprendre el funcionament del receptor per a acetilcolina pot implantar en un òvul els gens dels diferents components o els gens manipulats i observar el comportament que resulta; pot, en definitiva, preparar vacunes, hormones o proteïnes del plasma en quantitat i a baix preu per a usar-les en clínica.

Si hi ha algun camp en particular en què els canvis han sigut més inesperats ha sigut en el del coneixement dels gens eucariòtics, que fins a l'aparició del clonatge molecular estava a l'abric de les tècniques d'investigació. Una vegada que van començar a aïllar i estudiar alguns gens de virus eucariòtics primer i gens de cèl·lules eucariòtiques després es va poder constatar amb gran sorpresa que els gens eucariòtics no són continus com es pensava i com són els dels bacteris, sinó discontinus, i els trams que codifiquen o exons estan interromputs per trams intercalats o introns que no ho fan. L'existència d'eixos trams no codificadors serveix per a explicar, en part, l'existència d'un gran excés d'ADN en els organismes eucariòtics per damunt del que és necessari per a codificar la informació genètica. Però, sobretot, hi ha la possibilitat que pugui explicar en gran mesura l'evolució accelerada d'aquest tipus d'organització cel·lular. És fàcil que els introns hagen simplificat la recombinació i també poden haver permès que la construcció de noves proteïnes haja ocorregut barrejant les parts fins a aconseguir la combinació millor adaptada a les necessitats de la cèl·lula. En qualsevol cas, resulta interessant preguntar-se si els introns ja van existir en els primers organismes i van ser els eubacteris els únics a renunciar-hi després que havien complert un brillant objectiu en nom d'un estalvi de recursos, mentre la cèl·lula antecessora de què provenen els eucariotes va preferir conservar-los (Blake, 1983).

Abans que fóra coneguda l'estructura gènica dels eucariotes per a comparar-la amb la dels procariotes, semblava de sentit comú suposar que aquests últims són més primitius que els primers, perquè les cèl·lules procariòtiques són menys complexes i els seus genomes més petits. Però els descobriments dels últims anys han canviat tot això. Ja se sabia que la cèl·lula eucariòtica ha tingut un origen mixt. D'acord amb la teoria de l'endosimbiosi, els orgànuls, com ara mitocondris i cloroplasts, procedeixen de bacteris atrapats per una cèl·lula ancestral que, en comptes de destruir-les, va iniciar una convivència de profit mutu. Però de quin tipus de cèl·lula provenen el nucli i el citoplasma? Sens dubte, degué tenir organització procariòtica per definició. Però, de quin llinatge prové? Actualment se succeixen les respostes sobre això. La primera va ser que eubacteris i arqueobacteris es van separar al mateix temps que ho va fer la cèl·lula de la qual s'originarien el nucli i el citoplasma. Després va ser la que les cèl·lules eucariòtiques provenen d'arqueobacteris pel nucli i citoplasma i d'eubacteris pels orgànuls. Fins i tot s'ha proposat que els organismes originals van ser els arqueobacteris, de la branca de metanògens i halòfils extrems dels quals podien provenir els eubacteris, mentre nucli i citoplasma podrien haver-ho fet de la branca dels arqueobacteris que depenen del sofre. No coneixem encara la resposta, però sens dubte està tancada en les seqüències d'ADN i l'anàlisi ens revelarà en últim terme d'on venim. Però, i allò que som?

SEQÜENCIACIÓ DE L'ADN

Des de fa poc més d'un any hi ha hagut reunions als Estats Units quasi mensualment per a discutir sobre el projecte de dur a terme la seqüenciació completa del genoma humà.

Una de les primeres ocasions en què va ser tractada seriosament la qüestió va ser l'estiu de l'any passat en la conferència Gordon sobre genètica molecular. Quasi al mateix temps, Robert Sinsheimer, de la Universitat de Califòrnia a Santa Cruz, convocava una reunió informal sobre el mateix tema. La conclusió en els dos casos va ser essencialment la mateixa. El projecte, encara que gegantesc quant a la magnitud, era tècnicament factible.

Després prengué la iniciativa el Departament d'Energia dels Estats Units, que ha gastat bastants diners en estudis dels efectes de l'energia sobre la genètica humana (2.000 milions de dòlars en els últims anys). Charles Lisi, el nou director de l'Oficina d'Investigació sobre Salut i Ambient del Departament d'Energia, va patrocinar al març una aula taller a Santa Fe sota la presidència de Frank Ruddle, de la Universitat de Yale. En aquella reunió la discussió ja no va ser si pot ser seqüenciat tot el genoma humà, sinó com fer-ho. Es va considerar que el treball d'organització és tan important com la mateixa tecnologia de seqüenciació. Per això es va aconsellar fer un esforç no solament per aconseguir mètodes de seqüenciació més simples, més ràpids i menys costosos, sinó també i molt especialment en el desenvolupament de noves tècniques per a l'ús de dades. Per a aquests propòsits l'Oficina d'Investigació sobre Salut i Ambient podria invertir uns 20 milions de dòlars en els pròxims tres anys.

En aquella aula taller Walter Gilbert va dir amb entusiasme que “la seqüència completa del genoma humà constitueix el Sant Greal de la genètica humana” (Lewin, 1986). Fa sis anys, quan se li va concedir el premi Nobel, ja va afirmar que “les seqüències de l'ADN són les estructures definitives de la biologia molecular. No hi ha res més primitiu. Les preguntes es formulen allí en últim terme” (Kolata, 1980).

Walter Gilbert va compartir amb Paul Berg i Frederick Sanger el premi Nobel de química l'any 1980. D'acord amb la Reial Acadèmia sueca, el premi va ser concedit a Berg “pels seus estudis fonamentals sobre la bioquímica dels àcids nucleics, amb una referència particular a l'ADN recombinant. Berg va ser el primer investigador que va construir una molècula d'ADN recombinant, és a dir, una molècula que conté parts d'ADN de diferents espècies. El seu experiment pioner ha conduït a una nova tecnologia que se sol anomenar enginyeria genètica”. Sanger i Gilbert van ser guardonats pels descobriments de nous mètodes per a seqüenciar ADN. Aquests mètodes van fer que la seqüenciació de l'ADN passara a ser més fàcil i molt més exacta que la seqüenciació de proteïnes.

L'any 1974 Winston A. Salser deia en una revisió sobre les tècniques de seqüenciació d'ADN (Salser, 1974) que si diversos laboratoris uniren els esforços potser es podria seqüenciar el genoma de l'SV40. Encara que ja apuntava en eixa revisió la revolució que s'acostava amb l'ús del clonatge molecular, assenyalava com impossible la seqüenciació del genoma humà. Tenint en compte que una cèl·lula humana diploide conté uns 5pg d'ADN, és a dir, aproximadament un milió de vegades més ADN que el fag SV40, la grandària del genoma fa impràctic purificar amb un rendiment del 100%, la quantitat d'ADN de partida necessària per a proporcionar uns pocs micrograms d'un fragment de 1.000 pb seria enorme. No obstant això, les coses han canviat molt en l'última dècada.

Només fa nou anys que es va publicar en la revista *Nature* la primera seqüència d'un ADN humà (va ser la del gen que codifica la somatotropina coriònica) i ara no passa una setmana que se'n descriu alguna. Des d'aleshores, en vora 2.500 articles s'han assenyalat unes 500 seqüències humanes diferents, més molts altres fragments. Es tracta d'un augment exponencial en la producció científica sobre un tema concret, com ha esdevingut en molts altres casos anteriorment. Una cosa semblant va ocórrer, per exemple, quan es va descobrir l'existència i el paper fisiològic del camp a final de la dècada dels cinquanta. En totes les ocasions el creixement remet després de transcorregut un temps. No pot ser d'altra manera. S'ha calculat que si continuara el ritme actual de creixement en la

publicació de seqüències humanes per a l'any 1992 totes les publicacions biològiques s'ocuparien només d'eixe tema i, per a l'any 1994, totes les publicacions científiques farien el mateix.

Malauradament, en l'actual allau de publicacions sobre seqüències humanes hi ha redundància en un grau excessiu. S'ha descrit vuit vegades el clonatge de l'adenosina-desaminada i de les cadenes alfa i beta del receptor de les cèl·lules T. Encara que en alguns casos es tracte de formes al·lèliques, s'ha publicat 17 vegades la seqüència de l'oncogen c-myc i 33 vegades la seqüència del gen per a beta-globina. Avui publiquen aquestes seqüències unes 8 revistes científiques. Cal esperar que d'ara endavant eixes seqüències vagen directament a bancs de dades.

Carl Anderson, del Laboratori Nacional Brookhaven, va ser president de sessió d'una aula taller organitzada al març de 1979 per la Universitat Rockefeller. S'hi va tractar per primera vegada la necessitat de crear un banc de dades per a seqüències d'ADN (*DNA database*). El projecte es va perfilar més en una reunió de l'Institut Nacional per a les Ciències Mèdiques Generals de l'Institut Nacional de Salut dels Estats Units, que va tenir lloc al juliol de 1980. Però la base de dades d'ADN no va començar a funcionar fins a l'any 1982, quan ja estaven en marxa projectes semblants a Europa (Nucleotide Sequence Data Library del Laboratori Europeu de Biologia Molecular a Heidelberg) i al Japó. El contracte per a la base de dades d'ADN no el va guanyar la proposta presentada per Margaret Dayhoff de l'Institut d'Investigació Biomèdica de Washington DC, que té una gran experiència en l'emmagatzematge i la manipulació de seqüències de proteïnes. En canvi, el va obtenir la companyia Bolt, Beranek i Newman amb seu a Cambridge (Massachusetts), que té, alhora, contracte amb la companyia Los Alamos National Laboratory, amb Walter Goad com a principal investigador. Los Alamos tenia interès des de feia molt de temps en el tema. Ja l'any 1960 va fer reunions i conferències sobre la qüestió de la comparació de seqüències. El contracte per al projecte que s'anomena GenBank fou per 3 milions de dòlars, amb una durada de 3 anys.

GenBank tenia l'any 1982 seqüències amb 600.000 residus, la qual cosa comportava els 2/3 dels disponibles aleshores. El 1985 tenia seqüències amb 3 milions de residus. El 1986, aquest banc de dades, com també l'europeu, ha multiplicat per 25 els residus emmagatzemats en el primer any de treball. Malgrat tot, s'han generat retards molt superiors dels previstos. GenBank només té ara el 19% de les seqüències obtingudes l'any 1985. No obstant això, s'espera que el 1987, que és quan venç el contracte, hagen duplicat el que avui tenen emmagatzemat. Aquesta immensa tasca només es podrà aconseguir amb mesures dràstiques de simplificació en l'elaboració de la informació, com també mitjançant la introducció de prioritats per a determinats temes, com s'esdevé amb les investigacions sobre la SIDA.

En els bancs de dades una qüestió és l'emmagatzematge i l'oferta als investigadors de les seqüències emmagatzemades i una altra diferent la manipulació de la informació que contenen.

Les seqüències de l'ADN contenen diversos tipus d'informació. En primer lloc, informació sobre l'estructura, no tant com en el cas de l'ARN, però molta més de la que es pensava. L'estructura detallada de la doble hèlix depèn localment de quina siga la seqüència. En segon lloc, informació sobre l'estructura d'altres macromolècules: ARN i proteïnes. A més, informació sobre la manipulació per enzims i altres proteïnes de l'estructura i la informació continguda en l'ADN. Per

últim, també tenen molta informació sobre el camí seguit en l'evolució de les diverses seqüències i, fins i tot, sobre els mecanismes moleculars dels quals ha depès eixe procés. Aquests diferents tipus d'informació estan codificats d'acord amb claus diferents. Cal entendre que no hi ha només un codi genètic (terme que sol ser aplicat a la correspondència que hi ha entre nucleòtids en el missatger i aminoàcids a la proteïna, la síntesi de la qual controla el missatger), sinó diversos aplicables als diferents tipus de senyal per a manipular la informació. Aquests codis tenen un ampli camp d'aplicabilitat, però no són universals i presenten variants per a diversos tipus de sistemes i espècies.

També per a usar les seqüències d'ADN s'ha sentit la necessitat de crear institucions centralitzades. El 1984 l'Institut Nacional de Salut va concedir un contracte per 5 anys i 5,6 milions de dòlars a una petita companyia de Palo Alto (Califòrnia), IntelliGenetics, per a instal·lar un recurs nacional d'ordinadors per a biologia molecular. La companyia va sorgir el 1980 d'un projecte de col·laboració entre diversos departaments de Stanford que havia començat el 1975 amb el nom de MOLGEN. Es tractava de l'aplicació de la metodologia de la intel·ligència artificial a la biologia molecular. El recurs, que s'anomena BIONET, ha de proporcionar accés fàcil a bancs de dades de seqüències d'ADN i proteïnes, una biblioteca de programari sofisticat per a la recerca, el contrast i la manipulació de seqüències i, a més i principalment, ha de desenvolupar més programari. A més, un dels objectius més importants és contribuir a establir una comunitat de biòlegs moleculars que es comuniquen constantment i amb rapidesa (Lewin, 1984).

Les primeres seqüències d'ADN es publicaven en les revistes científiques, però quan van assolir una grandària excessiva es va plantejar la necessitat de suprimir eixa publicació, de manera que ara van directament a l'emmagatzematge en el banc de dades i els investigadors compleixen determinades formalitats en la presentació de la seqüència per a estalviar treball posterior d'aquesta manera.

La seqüència de l'ADN del fag IX174 de 5.375 residus va ocupar 2½ pàgines en la revista *Nature* (Sanger *et al.*, 1977). La seqüència de l'ADN del fag de T7 de 39.936 residus va ocupar 8 pàgines de la revista *Journal of Molecular Biology* (Dunn i Studier, 1983). La seqüència de l'ADN del genoma humà amb 3.000 milions de residus ocuparia 600.000 pàgines, és a dir, uns 400 llibres de 1.500 pàgines. És evident que les seqüències llargues han d'anar directament als bancs de dades. Així ha ocorregut amb la seqüència de l'ADN del virus d'Epstein-Barr de 172.282 residus, que no va ser publicat (hauria ocupat 34 pàgines d'una revista) sinó que es va dipositar directament en la base de dades del Laboratori Europeu de Biologia Molecular (Baer *et al.*, 1984).

SEQÜENCIACIÓ DE GENOMES

La proposta de seqüenciar en la totalitat el genoma humà no manca de precedents. Ja s'han aconseguit seqüenciar altres genomes i d'eixa seqüenciació s'han obtingut dades molt importants.

En un sentit estricte, el genoma és un conjunt de gens que tenen capacitat de replicació i expressió autònoma i que tenen una evolució pròpia. L'autonomia implica autosuficiència. Implica disposar de tots els components necessaris per a la síntesi de proteïnes. El genoma ha de contenir gens estructurals, que codifiquen proteïnes i els tres tipus d'ARN: missatger, ribosòmic i de transferència. Però l'autonomia no comporta independència total perquè

l'expressió de la informació sempre està subjecta a influències externes. A més de gens, el genoma ha de tenir múltiples senyals, que interpreten i obeeixen un conjunt de proteïnes. L'evolució opera sobre els gens i sobre els senyals i tracta d'adaptar l'organisme de la millor manera possible a les condicions del medi en què desenvolupa l'activitat vital. Però el terme genoma s'aplica també en un sentit ampli per a indicar un conjunt de gens la replicació del qual és autònoma, però l'extensió no ho és i als quals no manca evolució pròpia. És així com es parla de genomes de plasmidis, virus o orgànuls.

L'organització cel·lular eucariòtica, que és la que tenen els organismes superiors, es caracteritza, entre altres coses, per tenir orgànuls com mitocòndries i cloroplasts. Com a vestigi de la vida independent que un dia tingueren, els orgànuls de les cèl·lules eucariòtiques tenen el seu ADN, que codifica una sèrie de components del mateix orgànu i disposen dels mitjans necessaris per a la replicació del seu ADN i per a la síntesi de proteïnes codificades.

En aquests últims anys ha tingut lloc un considerable avanç en la seqüenciació de genomes d'orgànuls. Es coneix la seqüència nucleotídica completa de l'ADN mitocondrial d'alguns organismes. La seqüència humana amb 16.569 residus va ser la primera (Anderson *et al.*, 1981). Però avui es coneixen les de bestiar, ratolí i rata, com també les de mosca del vinagre, granota, *Trypanosoma*, *Leishmania*, gran part de la seqüència d'eriçó de mar, llevat i fongs filamentosos.

S'ha pogut comprovar que, encara que les funcions genètiques essencials de l'ADN mitocondrial s'ha conservat en múltiples organismes de diferent nivell evolutiu, l'estructura i l'organització dels gens i la forma d'expressar-se han evolucionat de diferent manera. Així, mentre el genoma mitocondrial humà i, en general, de mamífer mostra una estructura molt compacta amb els gens contigus i que manquen d'introns, els gens mitocondrials de llevat estan separats entre si per regions d'ADN no informatiu, algunes fins i tot fragmentades per introns. L'ADN mitocondrial humà codifica dos ARN ribosòmics, 22 tRNA i 13 proteïnes, totes integrades en sistemes de membrana del mateix orgànu. Sembla que són les mateixes proteïnes les que estan també codificades a les mitocòndries de les altres espècies.

La seqüenciació de l'ADN mitocondrial i la seqüenciació directa de proteïnes codificades van permetre establir la clau genètica utilitzada en l'orgànu. Així va sorgir la sorpresa. El codi genètic que es considerava que era universal va resultar no ser-ho. En les mitocòndries s'utilitzen variants diferents d'eixe codi. Després, per procediments anàlegs, s'ha pogut establir que hi ha també variants en alguns protozous i, almenys, en un bacteri.

La genètica de les mitocòndries va ser sempre quasi impossible d'analitzar perquè era molt difícil contactar amb mutants, ja que sempre resultaven letals. Tenir la seqüència ha permès entrar per a les mitocòndries en l'era de la genètica per anàlisi de l'ADN. Ara la major part de descobriments genètics no es fan amb la genètica tradicional, sinó amb tècniques bioquímiques. Avui dia el principal instrument és la hibridació molecular dels àcids nucleics basada en la complementarietat de les bases. Avui els mètodes d'hibridació quantitativa són, per als biòlegs moleculars, el mateix que han sigut els assajos d'activitat per als bioquímics de proteïnes. Essencialment, qualsevol gen per al qual hi haja una sonda d'hibridació pot ser clonat fins a obtenir-lo en quantitats de mil·ligrams amb què és molt fàcil la seqüenciació. A més, la genètica per anàlisi de l'ADN s'ha

completat després que s'haja aconseguit sistemes d'assaig fiables en què qualsevol gen aïllat és expressat amb exactitud.

En la carrera cap a la seqüenciació de genomes cada vegada més complicats els èxits més destacats són els aconseguits molt recentment per dos grups de japonesos (Gray, 1986). Ohyama *et al.*, de la Universitat de Kyoto, han seqüenciat el genoma del cloroplast de l'hepàtica *Marchantia polymorpha*, un briòfit. Es tracta d'un ADN amb 121.024 residus. D'altra banda, Shinozaki *et al.* han seqüenciat l'ADN del cloroplast de la planta de tabac, un traqueòfit. Eixe ADN en té 155.844. Els briòfits i els traqueòfits es van separar en l'evolució fa uns 400 milions d'anys, però els dos genomes tenen una organització molt semblant. La diferència de grandària no afecta pràcticament la informació que conté. Els dos genomes tenen uns 120 gens (gens per a ARN ribosòmics, gens per a 30 o 31 tRNA, gens per a 55 proteïnes identificades i altres 30 gens per a proteïnes encara no identificades). Entre les proteïnes identificades n'hi ha 19 que són ribosòmiques.

L'organització del genoma de cloroplast té clares semblances amb els genomes bacterians. Els gens amb relació funcional, com aquells, apareixen arraïmats. També la disposició dels gens per a ARN ribosòmics és la mateixa que en els bacteris. No obstant això, els gens dels cloroplasts presenten introns com els gens eucariòtics.

SEQÜENCIACIÓ DEL GENOMA HUMÀ

L'excitació generada per aquests èxits espectaculars en seqüenciació de genomes d'òrganul és un dels factors que més ha contribuït a la idea de tractar la seqüenciació completa del genoma humà. Però és convenient reflexionar sobre la magnitud del projecte.

La taula I mostra comparativament dades sobre els genomes de bacteris, fongs i mamífers, que tenen grandàries creixents. Com és fàcil d'entendre, quan es preparen fragments de l'ADN d'un genoma per a clonar-los, el nombre de fragments és menor com més gran siga la dimensió que tinguen. Però, tot i utilitzar els fragments més grans que avui es poden clonar, que són els d'uns 4.000 pb, seria necessari clonar uns 75.000 fragments per a poder seqüenciar el genoma humà complet. La xifra és enorme, però probablement caldria multiplicar-la per un factor que pot estar entre 3 i 10 si es vol treballar amb total garantia. Si es pren la grandària mitjana dels fragments com 50.000 pb, la taula II mostra els cromosomes d'una dona (hi apareixen els cromosomes sexuals X i Y) amb la longitud del seu ADN i el nombre de fragments que caldria clonar per a poder seqüenciar-lo. Com es veu, l'ADN més petit és el del cromosoma 21, que té 48 milions de parell de bases i l'estudi del qual requeriria clonar 2.900 fragments.

TAULA I

NOMBRE DE FRAGMENTES QUE CAL CLONAR

Grandària dels fragments clonats (pb)	Genoma bacteris (2×10^6 pb)	Genoma fongs (2×10^7 pb)	Genoma mamífers (3×10^9 pb)
5×10^3	400	4.000	600.000
10×10^3	200	2.000	300.000

20 x 10 ³	100	1.000	150.000
40 x 10 ³	50	500	75.000

Dahl *et al.*, 1981

TAULA II

CONTINGUT EN ADN DELS CROMOSOMES HUMANS; LONGITUD DE L'ADN DE CADA CROMOSOMA I NOMBRE DE CLONS NECESSARIS PERA CADA CROMOSOMA

<i>Cromosoma</i>	<i>Contingut en ADN (% del contingut en els autosomes)</i>	<i>Longitud de l'ADN (pb x 10⁶)</i>	<i>Nombre de clons (50 kb per clon)</i>
1	4,32	249	15.000
2	4,22	243	14.500
3	3,49	202	12.000
4	3,34	193	11.500
5	3,26	184	11.000
6	3,02	173	10.300
7	2,77	160	9.600
X	2,70	154	9.200
8	2,55	146	8.700
9	2,37	137	8.200
10	2,33	136	8.200
11	2,38	137	8.200
12	2,35	134	8.000
13	1,86	110	6.600
14	1,80	103	6.200
15	1,69	100	6.000
16	1,55	93	5.600
17	1,49	85	5.000
18	1,40	81	4.800
20	1,20	67	4.000
19	1,08	62	3.700
Y	0,92	53	3.200
22	0,86	52	3.100
23	0,82	48	2.900

Mendelsohn *et al.*, 1973

Actualment, una persona pot seqüenciar 100.000 nucleòtids a l'any al preu d'un dòlar per nucleòtid. Segons això, sembla que el projecte de seqüenciar el genoma humà amb la tecnologia actual costaria més de 2 mil milions de dòlars i requeriria uns 30.000 anys/persona. Al ritme actual de seqüenciació de 2 milions de nucleòtids per any requeriria entre mil i mil cinc-cents anys. El cost de l'ús de dades és de 3 cèntims per nucleòtid emmagatzemat per al volum de 6 milions de bases, encara que aquest pressupost s'ha vist desbordat perquè requereix personal més qualificat del que s'anticipava. En aquest cost, només l'ús de dades del projecte hauria de comportar uns 100 milions de dòlars en 10 anys. Per a comparar xifres de cost podem dir que la factura de col·lisionador de superconductivitat és de 3 mil milions de dòlars i de l'estació espacial és

de 9 mil milions de dòlars. A més, David Smith, del Departament d'Energia, opina que el projecte només implicaria 300 anys/persona, costaria 300 milions de dòlars i tardaria només 10-15 anys.

La reunió d'aquest estiu a Cold Spring Harbor va ser el simposi titulat Biologia Molecular de l'Homo Sapiens. Hi va haver molta discussió sobre si es tracta d'una aventura encertada escometre ara la seqüenciació del genoma humà (Lewin, 1986).

Tots els investigadors estan d'acord que seqüenciar el genoma humà és essencial. Com ha dit Sydney Brenner, director de la institució d'investigació mèdica Laboratori de Biologia Molecular, "qualsevol persona interessada en la investigació mèdica hauria d'estar interessada en la seqüència del genoma humà. Aquells que no ho estiguen, haurien de ser considerats com a subjectes que necessiten tractament mèdic". Paul Berg assenyalava que cal preguntar-se si el projecte val la pena, no en termes de dòlars, sinó pel que fa a l'impacte que hauria de tenir en el conjunt de la ciència biològica.

Al març d'aquest any Renato Dulbecco raonava en un comentari (Dulbecco, 1986) que si volem saber sobre el càncer ens hem de centrar en el genoma cel·lular. Segons ell, es presenten ara dues opcions: o tractar de descobrir els gens importants en càncer per un procediment fragmentari o seqüenciar el genoma complet d'una espècie seleccionada. Per a Dulbecco, no hi ha dubte que el mètode d'elecció és seqüenciar el genoma complet. Només així podrà haver-hi sondes per a tots els gens i podran ser classificats quant a l'expressió dels diversos tipus de cèl·lules. Aquesta classificació hauria de constituir la peça clau per a estudiar la progressió de la transformació maligna. Tampoc té cap dubte sobre l'espècie en què s'hauria de centrar l'esforç. Se sap ja que hi ha diferències en el control en espècies diferents, de manera que s'hauria de tractar la seqüenciació del genoma humà.

Però Dulbecco pensa que el coneixement del genoma humà no solament seria útil en l'estudi del càncer. Hauria de ser bàsic per al progrés de la fisiologia i la patologia humanes. Per exemple, el coneixement de la regulació de l'expressió de gens individuals en diferents tipus de cèl·lula hauria de beneficiar molts camps d'investigació, com ara l'estudi de desenvolupament i de l'organització del sistema nerviós. El coneixement adquirit es projectaria ràpidament en aplicacions terapèutiques molt diverses. L'esforç tècnic hauria de ser intens, però una millora de les tècniques actuals per un factor de 50 permetria completar el treball principal en 5 anys. També indicava que l'esforç havia de ser nacional i millor encara internacional, ja que "la seqüència de l'ADN humà és la realitat de la nostra espècie i tot el que passa al món depèn d'eixes seqüències".

Però la qüestió és si és encertat escometre el projecte ara. Sydney Brenner opina que és prematur. Eric Lander, de l'Institut Whitehead de Cambridge, creu que per a l'any 2000 d'una manera o altra tindrem la seqüència. Però el projecte en megaescala per a aconseguir-lo canviaria per a sempre l'estructura de la investigació biològica. David Bostein, del MIT, creu que el projecte posaria en perill a tots, però molt especialment els investigadors joves. Al seu parer, la primera cosa que fa falta és un mapa físic del genoma format per fragments de més de 40 kb que

s'encavalquen. David Baltimore opina de manera semblant i creu que el mapa costaria uns 10 milions de dòlars i tardaria a fer-se uns 3-5 anys.

Sembla que al Japó s'està considerant seriosament el projecte de seqüenciació de genoma humà. De fet, hi ha el projecte a l'Institut Riken de Tòquio dirigit per Eiichi Soeda. Creuen que seqüenciar el genoma humà seria important per a les millores en biotecnologia que se'n derivaran. Les noves tecnologies han fet la nova biologia molecular que ha accedit a àrees prohibides amb anterioritat al desenvolupament. Els nous desplegaments obriran encara més portes.

Akijoshi Wada, professor de biofísica de la Universitat de Tòquio, opina que la seqüenciació del genoma humà per 100 experts costaria 100 milions de dòlars i tardaria 250 anys. Però en una "fàbrica de seqüenciació" costaria 600 dòlars i requeriria només 30 anys. Wada, amb la col·laboració de la indústria i amb l'Agència Japonesa de Ciència i Tecnologia, està duent a terme des de 1981 un sistema automàtic per a la seqüenciació de l'ADN.

Seiko Instruments & Electronics ha desenvolupat ja un robot que fa la separació de fragments per als procediments de seqüenciació de Sanger i de Maxam-Gilbert. Toyo Soda Manufacturing Co. ha dissenyat un extractor específic per a ADN, amb una alta resolució de bases. Fuji Photo Film Co. ha creat un sistema per a la reproducció en massa de films per a l'electroforesi en gen. Finalment, Seiko i Hitachi Software Engineering Co. estan competint per a desplegar màquines que lliguen els gens i escriuen la seqüència.

Actualment, la Universitat de Tòquio i l'Institut Riken de Tòquio estan invertint 3-4 milions de dòlars en un esforç per a automatitzar la seqüenciació de l'ADN. Quan tots aquests avanços d'aquestes institucions i de la indústria s'unisquen, es preveu que serà possible seqüenciar 1 milió de bases al dia, de manera que es podrà escometre d'ací a dos anys la seqüenciació del cromosoma més petit del genoma humà, el cromosoma 21 amb 48 milions de bases per a ser completada en 5 mesos.

Malgrat tot, Nobuyoshi Shimizu, de la Universitat de Keio de Tòquio, afirma que no té el propòsit de seqüenciar tot el genoma humà. Per a aconseguir-ho cal una organització que, opina, el Japó no té.

Al juliol passat hi va haver una reunió de l'Institut Nacional de Salut a Bethesda patrocinada per l'Institut Mèdic Howard Hughes. El president de l'institut, Donald Fredricks, hi assenyala que es pot seqüenciar el genoma humà i es durà a terme el projecte, però la qüestió és com fer-ho, quan, per qui i també qui serà qui en pague la factura. Quasi ningú estava a favor de fer-hi l'esforç ara. La dificultat radica en els diners que costaria i la por que es retirara el finançament a altres temes d'investigació. Leroy va opinar que seria un error escometre el treball ara, perquè, sens dubte, d'ací a pocs anys s'haurà multiplicat l'actual velocitat de seqüenciació i haurà passat a ser un o dos ordres de magnitud més elevats que l'actual. De fet, Hood ha publicat ja la inscripció d'un mètode de seqüenciació d'ADN (Smith *et al.*, 1986). Apareixerà un seqüenciador automàtic que Hood ha contribuït a desplegar i que estarà prompte disponible comercialment.

Donald Friedrickson va dir que seria una gran equivocació embarcar-se en un esforç en gran escala per a la seqüenciació del genoma humà amb les tècniques d'estar per casa que ara tenim. En canvi, si es fan ara les inversions necessàries per a desenvolupar la tecnologia d'ací a 5 anys estarem en condicions de fer el treball amb molta més eficàcia i un cost molt inferior.

De manera que sembla que disminueix l'impuls per a seqüenciar el genoma humà i és substituït per un projecte de realització del mapa físic. Per a fer-ho cal fragmentar el genoma humà, separar fragments d'uns 40 kb, preparar una genoteca en cosmidis, que són els únics vectors capaços de dur fragment d'eixa dimensió i després ordenar-los en cadascun dels cromosomes. Tanmateix, s'entèn que després hi hauria altres dos fases. La segona consisteix en la seqüenciació de tot el genoma humà o part i la tercera implicaria la completa comprensió de la regulació de l'expressió de tots els gens humans.

MAPES GENÈTICS

Gran part del treball dels científics consisteix en l'elaboració de mapes. A mesura que els instruments d'exploració es fan més penetrants, revelen detalls cada vegada més petits de la regió sotmesa a escrutini.

Des de 1913 aproximadament se sap que els gens estan situats de manera monodimensional al llarg dels cromosomes. A més, se suposava ja aleshores que tots els individus de la mateixa espècie haurien de tenir els gens disposats en els cromosomes d'una manera molt semblant. Però era necessari establir la posició dels gens en els cromosomes. Primer, l'ordre en què estan. Després, el lloc que realment ocupen. Les dues informacions es presenten a través d'un mapa genètic. Els mapes genètics s'elaboren a partir de dades sobre les distàncies relacionades amb els gens i admeten, com ho fan els geògrafs, que les distàncies són additives. Per a fer cartografia genètica és necessari tenir marcadors, que se situen en el mapa com a referència.

Els mapes genètics clàssics estaven basats en la relació directa que hi ha entre la freqüència de recombinació dels gens i les distàncies relatives en el cromosoma. Per a preparar un mapa es necessitaven mutants, creus i caracterització d'una abundant progènia. Això ha funcionat molt bé amb bacteris i va ser dut als seus límits amb fags. Benzer va aconseguir revelar l'estructura detallada del gen. A final dels anys cinquanta l'anàlisi genètica era encara més potent que la microscòpia electrònica (o que la seqüenciació de l'ADN que encara no s'havia desenvolupat). Però amb eucariotes, com l'ésser humà, el mètode fallava perquè aquesta espècie té temps llargs de generació i molt petit nombre de cries.

Es va definir la genètica com la ciència de la variació. Fins que va arribar l'era de la genètica molecular no podia haver-hi genètica sense variació. Només es podia suposar l'existència d'un gen a través de l'existència de formes alternatives d'un caràcter genètic determinat. L'any 1977 ja es coneixien uns 1.200 *loci* genètics com a seu d'altres tants gens humans (100 al cromosoma X i 1.100 als autosomes). D'aquests, al voltant de 900 s'havien caracteritzat gràcies a l'existència de malalties congènites, la

majoria poc freqüents a causa de mutacions en els corresponents *loci* (McKusick i Ruddle, 1977). Però la genètica molecular va proporcionar altres mitjans per a inferir l'existència de gens específics. Una vegada coneguda l'existència d'un polipèptid es pot suposar l'existència del gen que el codifica, encara que no se'n coneguen variants.

En l'assignació dels gens a cromosomes específics, el desenvolupament de la genètica de les cèl·lules somàtiques va comportar un gran avanç. La realització de mapes amb aquest tipus de genètica necessita fraccionar el genoma, de manera que cromosomes o fragments cromosòmics puguin ser estudiats aïllats. La manera d'aconseguir-ho consisteix a usar cèl·lules de ratolí com a portadores del material genètic humà. Els cromosomes humans es transfereixen a cèl·lules de ratolí per fusió cel·lular, induïda per virus o substàncies químiques. Després se seleccionen les cèl·lules híbrides i es maten les cèl·lules parentals aprofitant deficiències que es compensen en les cèl·lules híbrides. Cada cèl·lula híbrida conserva uns cromosomes i en perd d'altres, i és relativament simple establir correlacions entre l'expressió de determinats gens i la presència d'un cromosoma. Així és com s'havia localitzat uns 120 gens en el genoma humà fa uns 4 anys (Tunnacliffe et al., 1984).

Hi ha dos tipus extrems de mapa genètic. Un mapa fenotípic en què una característica com una activitat enzimàtica o una malaltia congènita es localitzen en un cromosoma concret. I un mapa molecular en què s'expressa l'organització i la seqüència de l'ADN que compon els gens. Les distàncies que fan servir un tipus de mapa i l'altre difereixen unes 100 vegades.

Fer el mapa de seqüències d'ADN de fins a diversos milers de bases de llarg és ara un exercici de rutina, que combina tres tècniques bàsiques. Primer, el clonatge de fragments d'ADN, generalment en cèl·lules bacterianes. Després, l'electroforesi dels fragments i, finalment, la hibridació molecular. El clonatge de seqüències d'ADN serveix una varietat de propòsits en el context de l'elaboració de mapes d'ADN. Però la funció principal consisteix a descomposar el genoma de gran dimensió en fragments d'una longitud que en permeta l'anàlisi, com també l'obtenció en quantitat i en forma pura. El gran avantatge de basar la confecció de mapa en les seqüències clonades és que proporciona un mitjà de recuperar fragments definits del genoma en una forma que es pot utilitzar per a establir quin paper biològic exerceix (Southern, 1982). La preparació de mapes de centres de restricció determina la posició dels centres que reconeixen cada restrictasa al llarg de la cadena de l'ADN. Després, eixos centres es poden usar com a punts de referència per a mesurar les distàncies al llarg de la cadena d'ADN. La hibridació molecular d'ADN pot servir per a identificar seqüències específiques. La combinació de la hibridació molecular amb l'electroforesi en gel de fragments de restricció permet localitzar la posició de seqüències específiques en el mapa de restricció.

La distància entre una tècnica i l'altra s'aconsegueix cobrir mitjançant l'ús dels polimorfismes en la longitud dels fragments de restricció (polimorfisme de longitud dels fragments de restricció, RFLP en anglès). Són fàcils de detectar amb les tècniques normals de laboratori. Ja es coneixen diversos centenars d'aquests centres discrets de variació en el genoma humà.

UTILITAT DELS POLIMORFISMES

Els mapes genètics s'han fet seguint l'herència de la variació en caràcters concrets. Ara es pot mirar la variació de seqüència en segments concrets de l'ADN. Fa uns 10 anys que es va descobrir que les variacions de seqüència entre individus es podia detectar amb restrictases. Si es donava la circumstància que les variacions afectaven els centres que reconeixen els enzims, que d'eixa manera apareixien o desapareixien, s'alterava la grandària dels fragments que en resultaven de l'acció. Els polimorfismes en la longitud dels fragments obtinguts per l'acció de les restrictases es poden detectar amb sondes d'ADN clonat, que s'hibriden amb eixos fragments siga quin siga la grandària (que s'ha determinat independentment per electroforesi en gel). Si un RFLP s'hereta amb una pauta igual que una malaltia genètica és possible establir on resideix el gen de la malaltia, uns pocs milions de bases per dalt o per baix. I tot millora notablement si hi ha diversos marcadors. De fet, n'hi hauria prou amb un marcador polimòrfic per cada 20 milions de bases per a disposar d'un mapa adequat.

Jean Dausset està duent endavant un experiment cooperatiu poc corrent en el Centre per a l'Estudi del Polimorfisme Humà en el Collège de France a París. L'objectiu és l'elaboració d'un mapa detallat de marcadors genètics que comprega tots els cromosomes del genoma humà. Un mapa així permetria els investigadors encertar amb els gens defectuosos responsables de malalties hereditàries, fins i tot amb absència d'informació sobre els efectes bioquímics de les mutacions. Segons Victor McKusick, de la Facultat de Medicina de la Universitat John Hopkins, afecten l'espècie humana unes 1.300 malalties hereditàries. Per a la major part d'aquestes no s'ha aconseguit establir els efectes cel·lulars ni els gens implicats. En són exemples ben coneguts la fibrosi quística, la malaltia de Huntington, la distròfia muscular Duchenne i la malaltia granulomatosa crònica. Però una vegada que dispose del mapa de marcadors genètics, les famílies afectades per una malaltia congènita concreta podran veure's sotmeses a estudi per a saber si la malaltia s'hereta en conjunció amb alguns dels marcadors. Quan els dos marcadors s'hereten junts és probable que estiguen prop, en el mateix cromosoma. Així, el marcador serviria de referència per a localitzar el gen que causa el problema i, fins i tot, podria permetre'n l'aïllament i la caracterització. En el pitjor dels casos, el marcador es podria aprofitar per a diagnosticar la malaltia i per a identificar-ne els portadors. El mateix tipus d'aplicació podria tenir per a malalties que, com la diabetis o l'esquizofrènia, tenen, sens dubte, un component hereditari.

El projecte de Dausset pretén l'obtenció de suficients marcadors. Per a fer-ho s'han obtingut llinatges cel·lulars permanents de tots els membres de 40 famílies grans que comprenen diverses generacions identificades per investigadors d'arreu del món. Moltes les va aportar el grup de Raymond White en l'Institut Howard Hughes de la Facultat de Medicina de la Universitat d'Utah. Dausset ha proporcionat mostres a laboratoris de diversos països, que han de tenir una sonda d'ADN clonat per a un dels RFLP, l'herència del qual hauran de seguir en les 40 famílies. Dausset rep les dades i les compara amb ordinador amb les dades que fan referència a altres marcadors obtinguts per altres grups per a establir el grau de lligam entre els marcadors. Fins ara s'han usat uns 200 marcadors, encara que només una part tenen suficient polimorfisme per a ser útils.

Bona part del progrés recent en el diagnòstic de malalties hereditàries es basa en l'ús dels RFLP. L'avantatge del procediment és que permet detectar la presència d'una lesió hereditària encara que la mateixa lesió no s'haja caracteritzat (Newmark, 1984). Per a poder aprofitar la informació per al diagnòstic se'n necessita almenys un, i preferiblement diversos, que estiguen suficientment prop de la posició cromosòmica de la lesió perquè així siga mínima la probabilitat que els uns i els altres s'hagen separat per recombinació gènica en la formació dels gàmetes. Els polimorfismes no són resultat de la lesió, sinó que depenen de la variació preexistent en la població. Que un polimorfisme concret s'associe a una lesió en una família només depèn que existira en l'al·lel ancestral en què es va produir.

De vegades se sap on es troba la lesió i en què consisteix, com en el cas del diagnòstic prenatal i la detenció de portadors de fenilcetonúria clàssica en què hi ha el gen humà clonat per a l'enzim fenilalanina hidroxilasa (Woo *et al.*, 1983). Però en altres casos continua sent un misteri la naturalesa d'una lesió. Així s'esdevé amb la malaltia de Huntington, per a la qual Gusella *et al.* han descobert un marcador polimòrfic (Gusella *et al.*, 1983) per al tram en què aquesta té grandària suficient per a codificar fins a 25 proteïnes. En ocasions, la localització de la lesió en el mapa genètic condueix eventualment a la identificació de la seua naturalesa. El cas més interessant és el de la malaltia granulomatosa crònica (Royer-Pokora *et al.*, 1986). S'ha aconseguit clonar el gen que és anormal en l'afecció fagocítica en la forma lligada al cromosoma X. El clonatge s'ha fet sense referència a una proteïna específica i es basa, en canvi, en la posició en el mapa cromosòmic. La seqüència nucleotídica de l'ADN clonat prediu que ha de codificar una proteïna de 468 residus almenys, sense homologia amb cap proteïna coneguda i la funció de la qual, per tant, continua sent un misteri. No obstant això, presenta entre els residus 65 i 92 un tram hidrofòbic que podria constituir un domini transmembrana, de manera que probablement es tracte de proteïna integral de membrana. Allò que sí que està provat és que és defectuosa en quatre pacients de la malaltia.

De qualsevol manera, el diagnòstic per a persones que no se sap si són portadores o no de la malaltia ha de ser dut a terme amb el seu consentiment. Fa poc hi va haver una sol·licitud de fer l'assaig per a veure si portava la malaltia de Huntington una xiqueta de 2 mesos que seria adoptada i va ser rebutjada per la manca de consentiment. Avui hi ha ja marcadors per a un cert nombre d'afeccions hereditàries com ara distròfia muscular Duchenne, fibrosi quística, malaltia renal cíclica i s'estan buscant afanyadament marcadors per a malalties com la d'Alzheimer, la depressió i certs tipus de càncer. Però comença a fer-se assajos per a establir la presència de certes condicions a l'hora de considerar un candidat per a un lloc en la indústria. L'aplicació d'aquest cribatge genètic, en opinió de molts experts, constituirà un greu atemptat a la intimitat i a la igualtat d'oportunitats en l'ocupació. Hi ha ja un test, per exemple, per a detectar deficiències en alfa 1-antitripsina per a treballadors que estiguen en contacte amb amiant o pols de cotó.

Els polimorfismes de l'ADN han revolucionat l'anàlisi genètica humana, però els RFLP no es detecten amb freqüència i el grau de polimorfisme és baix. Està clar que l'anàlisi genètica se simplificaria molt si hi haguera sondes per a regions hipervariables del genoma humà, amb variació multial·lelica i altes heterozigosis. La primera regió així l'aïllaren per casualitat Wyman i White fa sis anys (Wyman i White, 1980). Després també per atzar s'han trobat

altres regions hipervariables associades al gen per a la insulina humana, gens per a alfa-globina i l'oncogen c-Ha-ras-1. En tots els casos la regió hipervariable consisteix en repeticions en tàndem d'una curta seqüència (minisatèl·lits). El polimorfisme resulta de diferències al·lèliques en el nombre de repeticions, que s'han d'originar per intercanvis desiguals en els processos de mitosi o meïosi per corriment de l'ADN durant la replicació. La variació que resulta en la longitud de minisatèl·lit pot ser detectada utilitzant qualsevol restrictasa que no talle la unitat repetida. Així s'aconsegueixen marcadors que s'hereten de manera estable.

El fragments obtinguts dels minisatèl·lits hipervariables són semblants als de RFLP, però les longituds varien no perquè s'hagen alterat els centres de reconeixement de restrictases per substitució d'un nuclètid per un altre, sinó per la variació en el nombre de seqüències repetides dins els fragments. En comparació dels RFLP convencionals que representen marcadors simples, aquests nous polimorfismes tenen localitzacions múltiples en què poden servir de marcador simultàniament. Finalment, resulten superiors als RFLP convencionals en què són molt més variables, de manera que moltes vegades la heterozigosi de la població s'acosta al 100% (Jeffrey *et al.*, 1985). A més a més, la tècnica de Jeffreys *et al.* té interessants aplicacions en medicina forense (Gill *et al.*, 1985).

ALGUNES CONSIDERACIONS FINALS

L'any 1963 Sydney Brenner va proposar a la institució d'investigació mèdica Laboratori de Biologia Molecular un projecte per a estudiar la biologia molecular del desenvolupament d'organisme pluricel·lular més senzill possible: el cuc d'1 mm de longitud i amb una vida de tres dies i mig *Caenorhabditis elegans*. Com deia en el projecte, "ens agradaria atacar el problema del desenvolupament cel·lular [...], elegir l'organisme diferenciat més senzill possible i sotmetre'l als mètodes analítics de la genètica microbiana". L'organisme consta de 959 cèl·lules somàtiques, de les quals 302 formen el sistema nerviós. Quan el treball va ser proposat, molts investigadors van pensar que estava avançat per a l'època. Watson va arribar a dir que d'ací a 20 anys. Però el projecte es va completar l'any 1983 (Lewin, 1984).

Ara es coneix el camí seguit en el desenvolupament per cadascuna de les cèl·lules de *C. elegans*, incloent-hi les connexions de les neurones en el sistema nerviós. I no sembla tenir cap pla senzill. Ni tan sols l'existència en el cos de l'animal d'una simetria bilateral es reflecteix en el desenvolupament, que sembla ser, en canvi, oportunista. S'han aconseguit un bon nombre de mutants que afecten el desenvolupament, amb els quals s'espera poder aprofundir en el coneixement del seu control.

Quan va acabar el projecte Brenner comentava que al començament s'havia dit que la resposta per a la comprensió del desenvolupament vindria d'un coneixement dels mecanismes moleculars del control genètic. Però els mecanismes moleculars són monòtons de tan simples i no ens diran el que volem saber. Hem d'intentar descobrir els principis d'organització. Com s'esdevé en gran quantitat de components, es munten en un mateix lloc, no crec que eixos principis d'organització estiguen incorporats en un dispositiu senzill com el codi genètic. En un principi es va pensar que el desenvolupament estaria imprès en el genoma com un programa semblant a la codificació de l'estructura de les proteïnes a través de la seqüència.

Però, segons Brenner, tendim a parlar de manera laxa sobre programes genètics, però hauríem de tenir cura amb les implicacions d'eixa expressió, fins i tot quan s'usa metafòricament. Ara prefereix usar termes com representació o descripció internes i diu que "l'explicació total de tot organisme resideix dins i un sent que ha d'haver-hi una gramàtica en alguna part de l'interior. En últim terme, l'organisme ha de ser explicable pel que fa als gens, simplement perquè l'evolució ha sigut a través d'alteracions en l'ADN. Però la representació no serà explícita. Cal entendre la gramàtica del desenvolupament per a poder trobar-hi sentit".

Brenner tracta d'aclarir el seu punt de vista utilitzant una analogia. El cap icosaèdric d'un fag és un objecte geomètric precís. Eixa geometria s'hereta, per la qual cosa hi ha d'haver una definició d'icosaedre en algun lloc del genoma. Però no hi ha cap gen que diga: fer un icosaedre. És diferent escriure un programa d'ordinador perquè dibuixi un icosaedre. Per a entendre allò que significa especificar un icosaedre cal entendre primer els principis d'assemblatge molecular, la manera en què les proteïnes de la càpsida interactuen i s'autoassemblen. L'icosaedre està codificat de manera que la informació està distribuïda per tot el genoma complet. És una bona imatge per a entendre què es necessita per a fer una mà o un fetge. Les especificacions d'aquestes estructures estan disperses per tot el genoma. No es tracta d'una informació clara i seqüencial com juntar aminoàcids per a fer una proteïna.

Tot això és molt pertinent amb el tema de la seqüenciació del genoma humà complet i què es podrà fer amb la informació que ens proporcione. No és que la possessió de la seqüència en si mateix dirà res. Però hauria de facilitar el camí cap a la comprensió definitiva del desenvolupament i les funcions de l'organisme humà. La seqüenciació es farà sens dubte i haurà de catalitzar un nou renaixement científic centrat en l'espècie humana, que passarà a ser així la millor coneguda i la que servisca de referència per al coneixement de les altres espècies animals. Projectes com el de la seqüenciació completa del genoma humà poden ser discutits en termes dels costos que representaran, com ha passat amb la investigació espacial, però és qüestió de temps i sempre més breu del que s'ha anticipat. En el cas de la biologia molecular, tots coincideixen que tanca la possibilitat de la comprensió dels sistemes biològics en el nivell bàsic, que és el molecular. Com ha dit Brenner, "la biologia molecular és l'art del que és inevitable, si hi treballes és inevitable que trobes com funciona, almenys al final".